Context- and cost-aware feature selection in ultra-low-power sensor interfaces

Komail Badami
MICAS @ ESAT,
KU Leuven, Belgium
komail.badami@esat.kuleuven.be

Wannes Meert
DTAI @ CS
KU Leuven, Belgium
wannes.meert@cs.kuleuven.be

Marian Verhelst
MICAS @ ESAT
KU Leuven, Belgium
marian.verhelst@esat.kuleuven.be

Abstract

This paper introduces the use of machine learning to drastically improve efficiency of ultra-low-power sensor interfaces. Adaptive feature extraction circuits are assisted by hardware embedded learning to dynamically activate only most relevant features. This selection is done in a *context and power consumption cost*-aware way, through modification of the C4.5 algorithm. As proof-of-principle, a Voice Activity Detector is expanded with the proposed context-and cost-dependent voice/noise classifier, resulting in an average circuit power efficiency gain of 80%, with negligible accuracy loss.

1 Introduction

Due to the increasing number of sensors integrated on a variety of ubiquitous electronic devices, the boundaries between the domains of machine learning, circuit design and signal processing are blurring more than ever [1]. Namely, extracting information from these always-on sensors is strongly constrained by the limited power availability in the sensing devices. To be able to deal with the sensory data overload in a power-efficient way, it is important to discard irrelevant data as soon as possible, and extract only information-carrying features. It has, for example, been well established that processing the raw data on-board is far more cost-effective (unless explicitly mentioned, in this paper cost refers to power consumption) compared to transmitting raw data to a central data collecting node [2]. Taking this one step further, also within a node relevant features should be extracted as close to the raw sensor as possible to avoid power wastage. Hence, adaptive ultra-low power consumption chips, such as our work in [3], demand an intelligent selection mechanism to pick essential features for information extraction, allowing the system to selectively enable and disable particular hardware feature extraction blocks. This paper will propose machine learning algorithms to enable such optimal hardware activation in future sensor interfaces.

In this regard, Section 2 will introduce the notion of context-aware feature activation, to dynamically activate current most relevant features. Section 3 further expands this idea to context-aware and feature cost-aware classification for improved resource efficiency, taking also feature circuit power cost into account. Section 4 applies the derived approach on a proof-of-principle hardware design of a Voice Activity Detector, demonstrating ~ 5X reduction in power consumption.

2 Context-aware feature selection

In many applications, the relative information content of a feature is highly context-dependent. Depending on the context, some features port a more distinctive value towards the classes of interest. Examples are acoustic classifiers, prone to various types of background noises, or patient-specific biomedical data classification. To always operate the target sensing systems at maximal resource efficiency, feature selection should be done in a context-aware way (See Fig. 1(a)). In

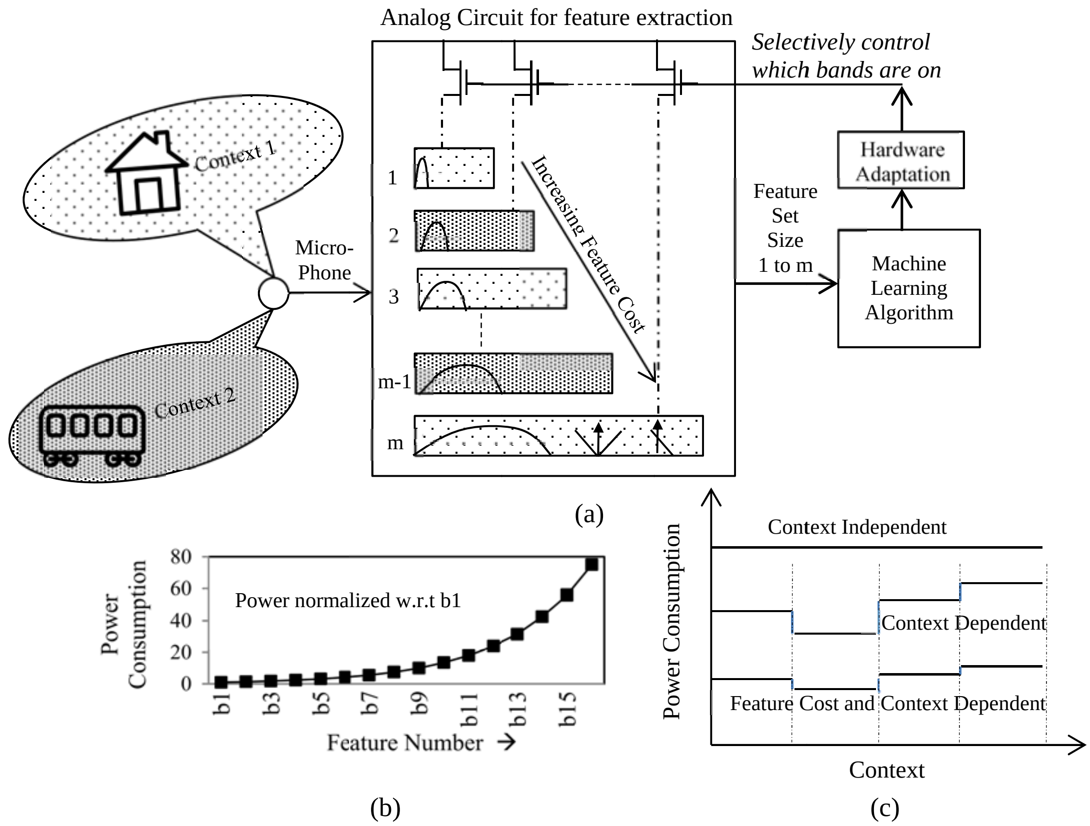

Fig. 1 (a) Top level schematic for Context- and Feature cost-aware classifier showing integration of adaptive hardware and machine learning; (b) illustrates the varying power cost across feature extraction hardware blocks; and (c) shows anticipated power consumption versus context for three types of classifiers. (All for the acoustic sensing scenario, introduced in Section 4.)

this work, decision tree based classifiers will be trained for different contexts with only discriminative features being selected per context. As will be demonstrated in Section 4, this context-awareness allows cutting the number of actively observed features by a factor of ~ 2X in typical sensing applications. To preserve power savings, the overhead of the introduced intelligence should be low. A tree based classifier is chosen for its amenability to low-footprint hardware implementation, comprising a series of comparisons. The decision tree automatically orders the features in order of decreasing information gain, allowing dynamically trading-off the classification accuracy with the power consumption.

3 Cost- and Context-aware feature selection

In many applications, the computational cost (c_i) to obtain a feature (f_i.) is not constant across available features as depicted in Fig 1(b). Previously introduced decision tree would learn the model only based on the information gain of the feature f_i, completely oblivious to the feature cost c_i. While state-of-the-art research, such as [4], introduced cost-aware learning for decision trees, cost optimization is primarily focused on misclassification cost, while the power impact of feature deactivation is not taken into account to the author's knowledge. Also, the algorithm proposed in [4] does not yet support continuous attributes and optimizing the traversal of the tree has negligible impact because the analog circuits have to be (de-)activated beforehand to capture sound.. Hence in very lower power applications it is of paramount importance to sensitize the machine learning algorithm to feature cost as well, enabling power consumption improvements as shown in Fig. 1(c). We suggest the following modification to C4.5 as in pseudo code in Fig 2. As such, instead selecting a feature with maximum Information-gain, the Information-gain/Watt is used as a metric. This allows the algorithm to relatively lower the rank of features with high feature cost. In other words, this modification gives "best value for money" or highest resource efficiency.

- If all labels are identical, create leaf node
- For each feature "f_i"
  - Find the normalized information-gain/watt from splitting on "f_i"
- Let "f_{best}" be the feature with the highest normalized information gain/watt
- If sufficient gain, create a decision node that splits on f_{best}, else create a leaf node
- Recurse on the sublists obtained by splitting on "f_{best}", and add those nodes as children of node

Fig. 2 Pseudo code depicting the power cost aware modification to C4.5

4 Proof-of-principle - Voice Activity Detector

This section will apply the proposed approach to the design of a relevant power-scarce sensor interface to illustrate its merits in a realistic design. An acoustic interface for Voice Activity Detection (VAD) in mobile devices, targeting continuous voice vs. noise classification, is extremely power-constrained due to its always-on operation. This design will serve as a proof-of-principle for both *context-aware* (section 4.1), respectively *context- and cost-aware* (section 4.2) voice/noise classification. Power savings of ~2X and ~5X resp. over *context independent* operation will be demonstrated.

Fig. 1(a) shows the implemented VAD system design, where an incoming voice/noise signal is processed in the analog domain by decomposing it into 16 frequency bands spaced on a logarithmic scale. Feature f_i is defined as mean signal level in each band over a time period 200ms. The actual feature cost is determined by executing analog circuit simulations of the filter bank, resulting in strongly rising feature power cost for higher frequency bands, as shown in Fig. 1(b).

All results reported in this paper have been conducted using the widely used NOIZEUS dataset [5] for voice and noise samples. Noise and voice samples are combined for varying signal-to-background-noise (SNR) values and context, using MATLAB. A balanced dataset of 90 samples each for voice and noise is used. Decision tree learning and evaluation is done deploying the modified C4.5 algorithm (J48 using Weka).

4.1 Context-aware Voice/Noise classifier

Within every NOISEUS context, a voice/noise classifier is learned in a supervised way. To ensure real life applicability, one common model is trained across a wide SNR-range, spanning from -20dB to +20dB. To avoid significant degradation in classification accuracy, relative instead of absolute signal levels should be used as basic features, being the difference between adjacent frequency band energy levels $\Delta f_i = f_{i+1} - f_i$. The differencing operation between the features incorporates SNR normalization, restoring accuracy. This leads to an efficient SNR independent model as illustrated with respect to street-noise/voice classification shown in Table. 1. The learnt context-dependent model shows that not all features are used in the classification and power is only spent on discriminative features. This can be seen in Table 2 where depending on the noise-context, only a subset of the features are used in voice/noise classification leading to ~2X savings in power consumption as compared to a context independent scenario where all the bands are used. Of course, context detection adds a cost penalty but contexts usually do-not change very often and is therefore not as expensive as voice recognition which runs continuously [6]. In this work we don't go into detail how the context is detected but for a cell phone this e.g. can be achieved by combining observed cellular IDs and inertial sensors [7].

4.2 Feature cost- and context-aware Voice/Noise classifier

Cost-awareness was introduced into J48 in Weka, to base classification on *Information-gain/Watt* rather than *Information-gain* alone as highlighted in Section 3. Rerunning model training, the impact of this modification can be seen in Table 2 where, as compared to context dependent scenario for the same context, relatively lower frequency bands are selected for voice/noise classification leading to ~5X power savings compared to the context independent case.

Table 1: Impact of wide SNR model training with feature f_i and with $\Delta f_i = f_{i+1} - f_i$

Classification Accuracy	Train and test @ SNR 0dB Feature f_i	Train @ SNR -20dB to 20dB Test @ SNR 0dB with Feature f_i	Train @ SNR -20dB to 20dB Test @ SNR 0dB with Feature $\Delta f_i = f_{i+1} - f_i$
Model Train	97.22	75	93
Model Test	95	70	86

Table 2: Selected features and power implications for Context / Cost (In)dependent scenarios

	Feature selection for context independent scenario																
Context	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**Power**
NA	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	**300**

	Feature selection for context dependent scenario																
Context	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**Power**
Street							■			■	■		■	■		■	186
Exhibition				■			■	■	■	■	■						57
Subway							■								■	■	137
Babble		■	■				■	■					■	■		■	165
Average																	**136**

	Feature selection for context and cost dependent scenario																
Context	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**Power**
Street		■	■	■			■	■	■	■	■						60
Exhibition		■	■	■			■	■		■	■						50
Subway					■		■		■						■		75
Babble		■	■				■	■	■	■				■			82
Average																	**67**

Table 3: Classification accuracy for Context / Cost (In)dependent scenarios

	Context Inde-pendent	**Context Dependent**					**Feature Cost and Context Dependent**				
		Street	**Exhib ition**	**Subway**	**Babble**	**Mean**	**Street**	**Exhib ition**	**Subway**	**Babble**	**Mean**
Training (-20dB to 20dB)	87	93	83	81	93	**87.5**	96	81	80	98	**88.75**
Testing @ -10dB	64	64	58	53	82	**64.25**	80	56	52	80	**67**
Testing @ 0dB	85	85	80	88	86	**84.75**	82	77	89	85	**83.25**
Testing @ 10dB	88	97	85	99	88	**92.25**	92	90	97	85	**91**
Size of tree	63	17	13	7	23	**15**	29	17	11	31	**22**
Power Cost	289	186	57	137	165	**136.25**	60	50	74	82	**66.5**

The analog circuit for feature extraction is designed such that the computation of un-necessary features can be deactivated. As shown in Table 3, there is no significant difference in the average accuracy numbers between *context aware* and *feature cost and context aware* to *context independent* classification. The downside is ~30% larger tree size for the feature cost-aware case but this is significantly outweighed by the power savings achieved.

5 Conclusions

This paper enables hardware-embedded machine learning to smartly control power-aware adaptivity in future sensor interfaces. Through the introduction of context- and cost-aware decision trees, hardware resource efficiency is maximized by controlled feature (de)activation. A VAD proof-of-principle demonstrates up to 5X power savings in real sensor interfaces, a gain invaluable towards ubiquitous sensing in the internet-of-things. We are currently working towards the chip tape-out of this self-adaptive context-aware voice activity detector.

References

[1] Hsu, Shu-Yu, et al. "A 48.6-to-105.2µW machine-learning assisted cardiac sensor SoC for mobile healthcare monitoring." *VLSI Circuits (VLSIC), 2013 Symposium on*. IEEE, 2013.

[2] Calhoun, Benton H., et al. "Body sensor networks: A holistic approach from silicon to users. Proceedings of the IEEE 100.1 (2012): 91-106.

[3] Verhelst, Marian, et al. "A reconfigurable, 0.13 µm CMOS 110pJ/pulse, fully integrated IR-UWB receiver for communication and sub-cm ranging." Solid-State Circuits Conference-Digest of Technical Papers, 2009. ISSCC 2009. IEEE International. IEEE, 2009.

[4] Ling, Charles X., et al. "Decision trees with minimal costs." Proceedings of the twenty-first international conference on Machine learning. ACM, 2004.

[5] Hu, Y., and P. Loizou. "NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms." (2005).

[6] Siewiorek, D., et al, "SenSay: A Context-Aware Mobile Phone," iswc, pp.248, Seventh IEEE International Symposium on Wearable Computers (ISWC'03), 2003

[7] Jiang, Yifei, et al. "User-Centric Indoor Air Quality Monitoring on Mobile Devices." AI Magazine 34.2 (2013): 11.